The ETA (Exercise Test Analyzer) expert system has been implemented and tested. The change in the health of the patient's heart, as measured by treadmill ECG tests, between any two tests was rated on a seven-point scale; each subject was rated on several features and overall. Rules for sub-area ratings were built from the verbal protocols of a POSCH cardiologist, and then weightings for combining sub-area ratings into an overall rating were determined. ETA was tested on 100 cases from the POSCH study and outperformed both the average POSCH cardiologist and a previously developed multiple regression model.

In the past year, the expert system ESCA ("Evaluator of Serial Coronary Angiograms") has been developed with domain knowledge organized in an inference network modeled after that of AGNESS. The domain knowledge was gathered from verbal protocols of a POSCH member inferring changes in atherosclerotic disease from changes in the flow of blood as revealed in angiograms taken at different times. In some cases, the POSCH member was first asked to determine the change solely from a form recording the consensus of a two-member sub-panel, and then was shown a more detailed and less stylized diagram and allowed to modify his conclusion. A sub-panel working from the films was also observed so the influence of the perceptual component could be judged. Indeed, much of ESCA's success is due to factoring the domain into a perceptual component followed by an expert system component. Its success thus dispels doubts about the applicability of expert system technology to domains with significant perceptual components. ESCA performed slightly better than the sub-panel of clinicians for the cases examined. Using ESCA for subjective clinical evaluation, and one cardiologist to screen the conclusions, POSCH can now evaluate films faster, more consistently, and with less cost.

Research in Progress

The research in progress for the current year will be a continuation of projects that have been underway for some time. The main areas will be --

1. *Inference engine mechanisms in diagnostic reasoning.* This will be a continuation of the Cleric/Vesalius project. The Cleric language will be used to model different diagnostic strategies -- path-following, compare and conquer, and stateless analysis.

2. *Merit system for question selection.* AGNESS is being used in developing an expert system for early detection of clinical trends in cystic fibrosis (CF) patients. In addition, the ESCA expert system will be extended to consider multiple lines of reasoning and to make use of the Dempster-Shafer method.

3. *Detection of deviations in time series by the human observer.* Surveillance and early detection of deviation from a homeostatic state are goals common to health care programs for the apparently healthy as well as for groups of patients known to have or have had specific diseases. Automated approaches to detecting deviations have the advantage of being reliably applied, traceable, consistent in outcome, and conserving of professional resources. Rule based expert systems based upon analysis of human graph reading strategies are being evaluated.

4. *Knowledge based system for improving transfusion practice.* The ESPRE expert system has undergone preliminary evaluation and is now being used in parallel with traditional decision processes in transfusion therapy.

D. List of Relevant Publications

1. Bailey, A. et al.: *Auditing, artificial intelligence, and expert systems,* DECISION SUPPORT SYSTEMS: THEORY AND APPLICATION, A.B. Whinston (ed.), North-Holland Publications, 1985.

2. Connelly, D.P. and Rich, S.S.: *Detection of deviations in time series by the human observer.* Proc. IMIA Workshop on Maintaining a Healthy State within the Individual. (in press)

3. Connelly, D. and Johnson, P.E.: *Medical problem solving.* Human Pathology, 11(5):412-419, 1980.

4. Elstein, A., Gorry, A., Johnson, P. and Kassirer, J.: *Proposed Research Efforts.* In D.C. Connelly, E. Benson and D. Burke (Eds.), CLINICAL DECISION MAKING AND LABORATORY USE. University of Minnesota Press, 1982, pp. 327-334.

5. Feltovich, P.J.: *Knowledge based components of expertise in medical diagnosis.* Learning Research and Development Center Technical Report PDS-2, University of Pittsburgh, September, 1981.

6. Feltovich, P.J., Johnson, P.E., Moller, J.H. and Swanson, D.B.: *The Role and Development of Medical Knowledge in Diagnostic Expertise.* In W. Clancey and E.H. Shortliffe (Eds.), READINGS IN MEDICAL AI, Addison-Wesley, 1984, pp. 275-319.

7. Johnson, P.E. and Zualkernan, I.: *Building expert systems for troubleshooting from knowledge of the design process,* IASTED International Conference on Expert Systems, Geneva, Switzerland, June, 1987.

8. Johnson, P.E., Zualkernan, I., and Garber,S.: *Specification of expertise,* International Journal of Man-Machine Studies (in press).

9. Johnson, P.E. *Cognitive models of expertise,* Proceedings of USC Symposium on Expert Systems and Auditor Judgment, February, 1986.

10. Johnson, P.E.: *Problem Solving.* In ENCYCLOPEDIA OF SCIENCE AND TECHNOLOGY, McGraw-Hill, 1985.

11. Johnson, P.E., Moen, J.B., and Thompson, W.B.: *Garden Path Errors in Medical Diagnosis.* In Bolc, L. and Coombs, M.J. (Eds.), COMPUTER EXPERT SYSTEMS, Springer-Verlag (in press).

12. Johnson, P.E.: *Cognitive Models of Medical Problem Solvers.* In D.C. Connelly, E. Benson, D. Burke (Eds.), CLINICAL DECISION MAKING AND LABORATORY USE. University of Minnesota Press, 1982, pp. 39-51.

13. Johnson, P.E.: *What kind of expert should a system be?* Journal of Medicine and Philosophy, 8:77-97, 1983.

14. Johnson, P.E.: *The Expert Mind: A New Challenge for the Information Scientist* In Th. M. A. Bemelmans (Ed.), INFORMATION SYSTEM DEVELOPMENT FOR ORGANIZATIONAL EFFECTIVENESS, Elsevier Science Publishers B. V. (North-Holland), 1984.

15. Johnson, P.E., Severance, D.G. and Feltovich, P.J.: *Design of decision support systems in medicine: Rationale and principles from the analysis of*

physician expertise. Proc. Twelfth Hawaii International Conference on System Science, Western Periodicals Co. 3:105-118, 1979.

16. Johnson, P.E., Duran, A., Hassebrock, F., Moller, J., Prietula, M., Feltovich, P. and Swanson, D.: *Expertise and error in diagnostic reasoning.* Cognitive Science 5:235-283, 1981.

17. Johnson, P.E. and Hassebrock, F.: *Validating Computer Simulation Models of Expert Reasoning.* In R. Trappl (Ed.), CYBERNETICS AND SYSTEMS RESEARCH. North-Holland Publishing Co., 1982.

18. Johnson, P.E. and Thompson, W.B.: *Strolling down the garden path: Detection and recovery from error in expert problem solving.* Proc. Seventh IJCAI, Vancouver, B.C., August, 1981, pp. 214-217.

19. Johnson, P.E., Hassebrock, F. and Moller, J.H.: *Multimethod study of clinical judgement.* Organizational Behavior and Human Performance 30:201-230, 1982.

20. Moller, J.H., Bass, G.M., Jr. and Johnson, P.E.: *New techniques in the construction of patient management problems.* Medical Education 15:150-153, 1981.

21. Sielaff, B.H., Galen, G., Scott, E. and Connelly, D.P.: *Knowledge-based system for improving transfusion practice.* In AAMSI Congress 87 (in press).

22. Simpson, D.E., Rich, E., Dalgaard, K., Gjerdingen, K., Crowson, T., O'Brien, D., Johnson, P.E.: *The diagnostic process in primary care: A comparison of general internists and family physicians.* SOCIAL SCIENCE AND MEDICINE (in press).

23. Slagle, J. et al.: "Expert systems in medical studies -- A new twist," *Proceedings of the Conference on Applications of Artificial Intelligence,* SPIE, 1986.

24. Slagle, J.: *An expert system for a resource allocation problem,* Communications of the ACM, September, 1985.

25. Slagle, J.: *Alpha Beta Pruning,* ENCYCLOPEDIA OF ARTIFICIAL INTELLIGENCE, S. Shapiro (ed.), John Wiley and Sons, Inc., New York, 1985.

26. Slagle, J. and Gaynor, M.: *An intelligent control strategy for computer consultation,* IEEE Trans. on Pattern Analysis and Machine Intelligence, March, 1984.

27. Spackman, K.A. and Connelly, D.P.: *Knowledge-based systems in laboratory medicine and pathology: A review and survey of the field.* Archives of Laboratory Medicine and Pathology 111:116-119, 1987.

28. Swanson, D.B., Feltovich, P.J. and Johnson, P.E.: *Psychological Analysis of Physician Expertise: Implications for The Design of Decision Support Systems.* In D.B. Shires and H. Wold (Eds.), MEDINFO77, North-Holland Publishing Co., Amsterdam, 1977, pp. 161-164.

29. Thompson, W.B., Johnson, P.E. and Moen, J.B.: *Recognition-based diagnostic reasoning.* Proc. Eighth IJCAI, Karlsruhe, West Germany, August, 1983.

E. Funding and Support

Work on the SOLVER project is currently supported by grants from the Control Data Corporation ($95,000; 1986-88) and IBM ($81,000; 1987) to Paul Johnson ($95,000; 1986-88) and by a grant from the Microelectronics and Information Sciences Center (MEIS) at the University of Minnesota to Paul Johnson, William Thompson, James Slagle ($300,000; 1986-7).

Research in medical informatics is supported, in part, by a training grant from the National Library of Medicine, LM-00160, in the amount of $712,573 for the period 1984-1989. Dr. Connelly and Prof. Johnson are participants in this grant. The post-doctoral fellowship of Dr. Spackman was funded by this grant.

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE

A. Medical Collaborations and Program Dissemination via SUMEX

Work in medical diagnosis is carried out with the cooperation of faculty and students in the University of Minnesota Medical School and St. Paul Ramsey Medical Center.

The Galen system is available on SUMEX from the University of Minnesota as an unsupported research tool for the study of recognition based reasoning systems.

B. Sharing and Interactions with Other SUMEX-AIM Projects

The SOLVER project has not been engaged in any formal sharing with other projects in the last year. The SUMEX resource has continued to serve as a communications vehicle for informal contacts with other researchers. Dr. Johnson conducted informal conferences during the year with Drs. Bruce Buchanan and William Clancey.

C. Critique of Resource Management

None.

III. RESEARCH PLANS

A. Project Goals and Plans

An overall goal of the project is to describe methods for the specification of expertise. Our objective is to construct an artifact (for example, an expert system) that can solve a class of problems which is currently solved by an expert. To construct this artifact a specification of the requirements is needed which outlines what needs to be computed to solve the problem.

A number of artifacts may achieve the same performance in a variety of ways. The expert's method works because it is adapted to the capabilities of the human information processing system and the demands of the problem-solving task. Since we may implement our specification on various kinds of processors, we seek a description that does not depend on a particular processing architecture. The purpose of knowledge acquisition is not to learn how to solve a problem, but rather to discover *what is required* to solve a problem.

Our goal is to use protocol records of problem-solving activity to develop a specification of the requirements for any artifact that would attempt to solve the same problem. Given a class of problems, such as medical diagnostic tasks, and a protocol record from experts solving these problems, the task is to determine a method for transforming the protocol into a specification of expertise.

Our goal is to investigate the following framework for specification of expertise:

1. The expert can be viewed as a processor that has the capability of producing certain problem-solving behavior using expertise. The task of knowledge acquisition is to determine this expertise.

2. The expert has developed a set of actions and abilities that are necessary to realize this expertise.

3. Although we cannot observe the expertise directly, we can observe the invocation of the expert's actions and abilities in a record of problem-solving behavior.

4. Since we can observe the invocation of actions and abilities by the expert, we can develop some representation of the expertise.

5. A statement of the expertise required to perform a task serves as a specification of the requirements for a computer program that is designed to perform the task.

The development of a specific methodology for collecting and analyzing protocol data to arrive at a formal specification of expertise.

B. Justification and Requirements for Continued SUMEX Use

Our current model development takes advantage of the sophisticated Lisp programming environments on SUMEX and local facilities. Although much current work with Galen is done using a version running on a local VAX 11/780, we continue to benefit from the interaction with other researchers facilitated by the SUMEX system. We expect to use SUMEX to allow other groups access to the Galen program. We also plan to continue use of the knowledge engineering tools available on SUMEX.

We have completed a CommonLisp implementation of the Galen system and expect to rely heavily on CommonLisp for future projects.

C. Needs and Plans for Other Computing Resources Beyond SUMEX-AIM

Our current research support has permitted us to purchase Sun workstations for our Artificial Intelligence laboratory. The availability of CommonLisp on these machines is one reason why we expect to make use of that language in the future.

SUMEX will continue to be used for collaborative activities and for program development requiring tools not available locally.

D. Recommendations for Future Community and Resource Development

As a remote site, we particularly appreciate the communications that the SUMEX facility provides our researchers with other members of the community. We, too, are moving toward a workstation-based development environment, but we hope that SUMEX will continue to serve as a focal point for the medical AI community. In addition to communication and sharing of programs, we are interested in development of CommonLisp based knowledge engineering tools. The continued existence of the SUMEX resource is very important to us.

IV.B.5. ATTENDING Project

ATTENDING Project--Expert Critiquing Systems

Perry L. Miller, M.D. Ph.D.
Department of Anesthesiology
Yale University School of Medicine
New Haven, CT 06510

I. SUMMARY OF RESEARCH PROGRAM

A. Project rationale

Our project is exploring the "critiquing" approach to bringing computer-based advice to the practicing physician.

Critiquing is a different approach to the design of artificial intelligence based expert systems. Most medical expert systems attempt to simulate a physician's decision-making process. As a result, they have the clinical effect of trying to tell a physician what to do: how to practice medicine. In contrast, a critiquing system first asks the physician how he contemplates approaching his patient's care, and then critiques that plan. In the critique, the system discusses any risks or benefits of the proposed approach, and of any other approaches which might be preferred. It is anticipated that the critiquing approach may be particularly well suited for domains, like medicine, where decisions involve a great deal of *subjective* judgment.

To date, several prototype critiquing systems have been developed in different medical domains:

1. ATTENDING, the first system to implement the critiquing approach, critiques anesthetic management.
2. HT-ATTENDING critiques the pharmacologic management of essential hypertension.
3. VQ-ATTENDING critiques aspects of ventilator management.
4. PHEO-ATTENDING critiques the laboratory and radiologic workup of a patient for a suspected pheichromocytoma.
5. In addition, a domain-independent system, ESSENTIAL-ATTENDING, has been developed to facilitate the implementation of critiquing systems in other domains.

C. Highlights of Research Progress

Current projects include the following:

HT-ATTENDING The original prototype version of HT-ATTENDING has been converted to the ESSENTIAL-ATTENDING format, and updated to reflect current thinking in the field of hypertension management. A major priority is to subject this system to validation and clinical evaluation, and to explore how best to disseminate the system as a practical consultation tool.

DxCON: Critiquing Radiologic Workup DxCON extends the design developed in PHEO-ATTENDING to critique the radiologic workup of suspected obstructive jaundice. Workup is an area in which we will aggressively pursue the critiquing approach for two reasons. 1) Since many areas of workup are quite constrained, it may prove possible to develop and test complete systems in a reasonably short time-frame. 2) Since workup is expensive, and very wasteful of resources if performed improperly, a computer system which helps to optimize a physician's workup plans could have significant economic benefits. The present national emphasis on controlling health costs makes this project very topical. We are also using this domain to explore issues of knowledge acquisition and verification.

ICON: Critiquing Radiological Differential Diagnosis Most existing diagnostic computer systems produce a ranked differential diagnosis as their output. In this process, the rich structure of the knowledge that went into developing the diagnoses may be lost to the user. ICON explores a different approach to diagnostic advice in the domain of radiology. To use ICON, a radiologist describes a set of findings seen on chest x-ray, together with a proposed diagnosis. ICON then produces a detailed analysis of *why* the observed findings serve to support or to rule out the diagnosis. It may also suggest further findings that might help refine the diagnosis, again explaining why the findings are important.

D. Publications

1. Miller, P.L.: Expert Critiquing Systems: Practice-Based Medical Consultation by Computer. New York: Springer-Verlag, 1986.

2. Miller, P.L. (Ed.): Selected Topics in Medical Artificial Intelligence. New York: Springer-Verlag (in press).

3. Miller, P.L., Shaw, C., Rose, J.R., Swett, H.A.: Critiquing the process of radiologic differential diagnosis. Computer Methods and Programs in Biomedicine 22:12-25, 1986.

4. Miller, P.L.: The evaluation of artificial intelligence systems in medicine. Computer Methods and Programs in Biomedicine 22:5-11, 1986.

5. Rennels, G.D., Shortliffe, E.H., Miller, P.L.: Choice of explanation in medical management: A multi-attribute model of artificial intelligence approaches. Medical Decision Making 7:22-31, 1987.

6. Rennels, G.D., Miller, P.L.: Artificial intelligence research in anesthesia and intensive care. Anesthesiology (submitted).

7. Miller, P.L., Rennels, G.D.: Prose generation from expert systems: An applied computational linguistics approach (submitted).

8. Mars, N.J.I., Miller, P.L.: Knowledge acquisition and verification tools for medical expert systems. Medical Decision Making 7:6-11, 1987.

9. Miller, P.L., Blumenfrucht, S.J., Rose, J.R., Rothschild, M., Swett, H.A., Weltin, G., Mars, N.J.I.: HYDRA: A knowledge acquisition tool for expert systems which critique medical workup. Medical Decision Making 7:12-21, 1987.

10. Swett, H.A., Miller, P.L.: ICON: A computer-based approach to differential diagnosis in radiology. Radiology (in press).

11. Rennels, G.D., Shortliffe, E.H., Stockdale, F.E., Miller, P.L.: A computational model of reasoning from the clinical literature. Computer Methods and Programs in Biomedicine (in press).

12. Rennels, G.D., Shortliffe, E.H., Stockdale, F.E., Miller, P.L.: A structured representation of the clinical literature and its use in a medical management advice system. Bulletin du Cancer (in press).

13. Miller, P.L., Barwick, K.W., Morrow, J.S., Powsner, S.M., Riely, C.A.: Semantic relationships and medical bibliographic retrieval: A preliminary assessment (submitted).

14. Miller, P.L.: Exploring the critiquing approach: Clinical practice-based feedback by computer. Biomedical Measurement, Informatics and Control (submitted).

15. Rennels, G.D., Shortliffe, E.H., Stockdale, F.E., Miller, P.L.: A computational model of reasoning from the clinical literature. The AI Magazine (accepted pending revision).

16. Miller, P.L.: Exploring the critiquing approach: Sophisticated practice-based feedback by computer. Proceedings of the Fifth World Conference on Medical Informatics MEDINFO-86, Washington, D.C., October 1986, pp 2-6.

17. Mars, N.J.I., Miller, P.L.: Tools for knowledge acquisition and verification in medicine. Proceedings of the Tenth Symposium on Computer Applications in Medical Care, Washington, D.C., October 1986, pp. 36-42.

18. Miller, P.L., Blumenfrucht, S.J., Rose, J.R., Rothschild, M., Weltin, G., Swett, H.A., Mars, N.J.I.: Expert system knowledge acquisition for domains of medical workup: An augmented transition network model. Proceedings of the Tenth Symposium on Computer Applications in Medical Care, Washington, D.C., October 1986, pp. 30-35.

19. Rennels, G.D., Shortliffe, E.H., Stockdale, F.E., Miller, P.L.: Reasoning from the clinical literature: The Roundsman system. Proceedings of the Fifth World Conference on Medical Informatics MEDINFO-86, Washington, D.C., October 1986, pp. 771-775.

20. Rennels, G.D., Shortliffe, E.H., Stockdale, F.E., Miller, P.L.: Updating an expert knowledge base as medical knowledge evolves: Examples from oncology management. Proceedings of the American Association of Medical Systems and Informatics Congress-87, San Francisco, May 1987, pp. 238-231.

21. Fisher, P.R., Miller, P.L., Swett, H.A.: A script-based representation of medical knowledge involving multiple perspectives. Proceedings of the American Association of Medical Systems and Informatics Congress-87, San Francisco, May 1987, pp. 233-237.

22. Miller, P.L.: Expert consultation systems in medicine: A complex and fascinating domain. Proceedings of the Annual Meeting of the IEEE (Electro-87), New York, April 1987, pp. 1/2:1-4 (invited paper).

23. Miller, P.L., Fisher, P.R.: Causal models in medical artificial intelligence. Proceedings of the Eleventh Symposium on Computer Applications in Medical Care, Washington, D.C., November 1987 (submitted).

24. Powsner, S.M., Barwick, K.W., Morrow, J.S., Riely, C.A., Miller, P.L.: Coding semantic relationships for medical bibliographic retrieval: A preliminary study. Proceedings of the Eleventh Symposium on Computer Applications in Medical Care, Washington, D.C., November 1987 (submitted).

E. Funding Support

EXPERT COMPUTER SYSTEMS WHICH CRITIQUE PHYSICIAN PLANS
NIH Grant R01 LM04336
Principal Investigator: Perry L. Miller, M.D., Ph.D.
Annual Direct Costs: approximately $100,000
Period of Support: 9/1/85-8/31/87

This two-year grant supports the exploration of the critiquing approach to bringing computer-based advice to the physician, focusing primarily on the underlying system design issues.

SUPPORT OF THE UNIFIED MEDICAL LANGUAGE PROGRAM
NLM Contract N01-LM-6-3524
Principal Investigator: Perry L. Miller, M.D., Ph.D.
Annual Direct Costs: approximately $100,000
Period of Support: 8/22/86-8/21/88

This two-year research contract is part of the NLM Unified Medical Language (UML) program. We are defining a set of semantic relationships which could be used to augment the UML, to facilitate such functions as medical bibliographic retrieval.

SUPPORT FOR MEDICAL INFORMATICS AND ARTIFICIAL INTELLIGENCE
Ira DeCamp Foundation
Co-Principal Investigators: Henry A. Swett, M.D.
Perry L. Miller, M.D., Ph.D.
Annual Costs: $75,000
Period of Support: 7/1/86-6/30/90

This grant supports our present Medical Informatics program and is currently being used primarily to support Medical Informatics research training. If the present training application is funded, the Ira DeCamp support could be used for other activities in support of the training such as for a program secretary and for computing programming support.

MEDICAL INFORMATICS RESEARCH TRAINING AT YALE
Principal Investigator: Perry L. Miller, M.D., Ph.D.

We have been informed that we will receive a five-year training grant starting July 1, 1987.

Pending Support

EXPERT COMPUTER SYSTEMS WHICH CRITIQUE PHYSICIAN PLANS
Principal Investigator: Perry L. Miller, M.D., Ph.D.
Annual Direct Costs: approximately $100,000
Period of Support: 9/1/87-8/31/90

This grant requests continuation of our currently funded grant which is exploring the critiquing approach to bringing computer-based advice to the practicing physician. This continuation grant application focuses especially on refining and evaluating the HT-ATTENDING system which critiques hypertension management.

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE

Until recently we have been using the RUTGERS-AIM Resource. We used that facility to implement all of our early critiquing systems. We are currently in the early stages of moving part of our critiquing research to the SUMEX-AIM facility. Our main uses of SUMEX-AIM will be the following:

1. We will use SUMEX-AIM to demonstrate two of our systems, ATTENDING and HT-ATTENDING.

2. We will use SUMEX-AIM for the continued refinement of HT-ATTENDING, and for a planned controlled clinical experiment to measure the effect of HT-ATTENDING's advice on patient care. This will be performed in the Yale New Haven Hospital Primary Care Center, and is planned to commence this coming year.

3. We will use SUMEX-AIM for communication access to the national AIM community.

We have found our use of the RUTGERS-AIM facility to be extremely valuable. It provided us the resources needed to initiate our research and to continue several projects which are still active. It provided a natural vehicle to allow us to demonstrate the various systems easily, both in the United States and in Europe. Also, it enabled us to collaborate very closely with Dr. Glenn Rennels in his Stanford Medical Information Science thesis project on the Roundsman system. Via SUMEX-AIM and RUTGERS-AIM, Dr. Rennels and Dr. Miller maintained very close contact, typically with multiple messages each week, and sometimes within a single day.

III. FUTURE PLANS

We plan to continue our critiquing research as outlined above. One of our highest priorities will be the controlled experimental evaluation of the HT-ATTENDING system, which will be done using SUMEX-AIM. We will also continue to utilize SUMEX-AIM as outlined above. Although we are increasingly moving a great deal of our work onto internal workstations, we nevertheless plan to continue our use of SUMEX-AIM, especially in the further refinement and evaluation of HT-ATTENDING.

IV.C. Pilot Stanford Projects

Following are descriptions of the informal pilot projects currently using the Stanford portion of the SUMEX-AIM resource, pending funding, full review, and authorization.

In addition to the progress reports presented here, abstracts for each project are submitted on a separate Scientific Subproject Form.

IV.C.1. REFEREE Project

REFEREE Project

Bruce G. Buchanan, Ph.D., Principal Investigator
Computer Science Department
Stanford University

Byron W. Brown, Ph.D., Co-Principal Investigator
Department of Medicine
Stanford University

Daniel E. Feldman, Ph.D., M.D., Associate Investigator
Department of Medicine
Stanford University

I. SUMMARY OF RESEARCH PROGRAM

A. Project Rationale

The goals of this project are related both to medical science and artificial intelligence: (a) use AI methods to allow the informed but non-expert reader of the medical literature to evaluate a randomized clinical trial, and (b) use the interpretation of the medical literature as a test problem for studies of knowledge acquisition and fusion of information from disparate sources. REFEREE and REVIEWER, a planned extension, will be used to evaluate the medical literature of clinical trials to determine the quality of a clinical trial, make judgements on the efficacy of the treatment proposed, and synthesize rules of clinical practice. The research is an initial step toward a more general goal - building computer systems to help the clinician and medical scientist read the medical literature more critically and more rapidly.

B. Medical Relevance

The explosive growth of the medical literature has created a severe information gap for the busy clinician. Most physicians can afford neither the time required to study all the pertinent journal articles in their field, nor the risk of ignoring potentially significant discoveries. The majority of clinicians, in fact, have little sophistication in epidemiology and statistics; they must nonetheless base their pragmatic decisions on a combination of clinical experience and published literature. The clinician's computerized assistant must ferret out useful maxims of clinical practice from the medical literature, pass judgment on the quality of medical reports, evaluate the efficacy of proposed treatments, and adjudicate the interpretation of conflicting and even contradictory studies.

C. Highlights of Progress

REFEREE, a rule-based system built upon the EMYCIN framework, partially encodes the epidemiological knowledge of two highly regarded experts at Stanford, a biostatistician (Dr. Bill Brown) and a clinician (Dr. Dan Feldman). The REFEREE system, in particular, allows the informed but non-expert reader of the medical literature to evaluate the believability of a randomized clinical trial.

In the future, REFEREE and its extensions will alleviate the knowledge-acquisition

bottleneck for an automated medical decision-maker: the program will evaluate the quality of a clinical trial, judge the efficacy of the treatment proposed therein, and synthesize rules of clinical practice. For the present, however, the fusion of knowledge from disparate sources remains a problem in pure AI. The efforts of the REFEREE team have instead focused their efforts on the refinement and deepening of REFEREE's biostatistical knowledge by applying effective knowledge acquisition and knowledge engineering techniques. Dr. Diana Forsythe and Dr. Harold Lehman are developing and using interview methods to acquire this knowledge from Dr. Brown, and R. Martin Chavez is implementing this in the prototype REFEREE expert system.

The REFEREE prototype is a consultant that evaluates the design and reporting of a single conclusion from randomized control trial for its believability. It contains, in preliminary form, Professor Brown's expert knowledge of biostatistics. REFEREE evaluates each statistical procedure described by the authors of the paper. The automated consultant then determines the most appropriate method for the problem at hand, based on the design of the trial and the hypotheses to be tested. REFEREE checks critical assumptions, looks for possible statistical abuses, verifies adjustments, and re-computes the statistics. In a beta-blocker study that employs the Cox proportional-hazards model, for instance, REFEREE will analyze the Kaplan-Meier survival curve and verify or reject the presence of a significant treatment effect.

The Knowledge Base: In order to evaluate the paper's presentation of a statistical test, REFEREE must apply three kinds of knowledge:

1. the statistical techniques that are relevant to the kinds of data likely to be found in a randomized clinical trial.

2. the methods to perform statistical tests to verify the paper's results.

3. the techniques to test hypotheses, to determine if the data in a paper support the conclusions of that paper.

Randomized controlled trials are used to test hypotheses regarding the effectiveness of various kinds of medical interventions. Dr. Brown classifies studies on the basis of three major attributes: the type of intervention tested (e.g. drug, surgery, health process change, etc.); the type of endpoint against which that intervention was tested (e.g. mortality, objective morbidity, subjective morbidity, etc.); and the type of conclusion drawn by the investigator/author on the basis of the research (e.g. that different treatments do or do not produce different outcomes, that a particular treatment is or is not cost-effective, etc.). Following this classificatory scheme, we decided to begin by producing a prototype REFEREE system that would help the reader to evaluate a single published conclusion concerning the effect of a given drug treatment on mortality.

Knowledge Acquisition: Having defined the scope of the initial knowledge base, we turned to the problem of collecting the information from Dr. Brown for inclusion in the system, i.e. knowledge acquisition. This task generally involves a relatively long-term process of face-to-face information gathering during sessions between the expert and one or more knowledge engineers. Dr. Diana Forsythe has noted a parallel between the communicative and analytical tasks involved in knowledge acquisition and those undertaken in ethnographic research. For this reason, we included an anthropologist in the research team and make use of ethnographic techniques in order to maximize the efficiency and quality of the data collection process.

Dr. Lehmann and Dr. Forsythe have carried out several months of systematic interviews with Dr. Brown in order to begin the process of constructing and refining the knowledge base for the current REFEREE prototype. We have combined a case-based approach that allows us actively to observe Dr. Brown as he reads papers, with semi-

directed interviewing oriented toward understanding his terminology and category system. We find that these techniques work very well: Dr. Brown's interest in the knowledge acquisition process has been sustained, and indeed has increased over time as the system based on his expertise has evolved. He is clearly comfortable with this approach, and notes that it has actually afforded him additional insight into the way he interprets the literature.

In order to codify the information gathered from Dr. Brown, Dr. Lehmann chose a model based on the *influence diagrams* used in decision analysis, in which the expert indicates which factors or *parameters* he finds crucial in making his judgement about the quality of the paper. Based on information from our expert, we have taken "believability" as the primary parameter of the present system, defined operationally by Dr. Brown as "the odds I am willing to give that the conclusions of the paper would be replicated in an experiment based on the methods reported in the paper but without any of the flaws". Within the influence diagram, parameters are connected to each other in a structure indicating the information considered by Dr. Brown in making particular judgments. In assessing believability, for instance, he considers the acceptability of the randomization, the quality of the blinding, other sources of bias, and how well the results substantiate the conclusion. Our use of influence diagrams has numerous advantages: the approach is acceptable to Dr. Brown, it is flexible, it can represent several aspects of the structure of the knowledge used by the expert, and the resultant data can be entered easily into the computer.

Once entered into the machine, the influence diagram is converted into *rules* such as the following:

```
If   : The quality of the randomization is high and
       The quality of the blinding is poor and
       The other sources of bias are unknown and
       The results substantiate the conclusion,

Then : There is suggestive evidence (0.7) that the believability of the
       clinical trial is high.
```

The number (0.7) captures the uncertainty of the expert in drawing a specific conclusion from the specific antecedents; this number is known as a *certainty factor*. The mathematics of certainty factors has been widely discussed in the literature.

Inference in REFEREE: REFEREE was originally built within EMYCIN, an AI environment developed from MYCIN at Stanford. In 1986 Chavez introduced some fundamental improvements to the REFEREE program; among other things, these changes greatly improved communication with the user (see "The User Interface", below).

The system is programmed to act as a problem solver, following the rules in the knowledge base in a *backwards chaining* path. For instance, the machine has the determination of the paper's believability as its *goal*. At the outset it finds a rule that reasons about the paper's believability (the above example). It then examines each antecedent of that rule in turn and looks for rules that draw a conclusion on *that* parameter, recursively, until an antecedent is found that has no rules. REFEREE then queries the user about that antecedent. For instance, from the rule "If the method of randomization was reported and the design of the randomization was good and the implementation of the randomization was poor - Then there is suggestive evidence (.6) that quality of the randomization method was acceptable", the machine would find that there are no rules that conclude that the method of randomization was reported. It would then ask the user, "Was the method of randomization reported?" If the answer is "No", then the machine abandons the rule in question, but saves the response for

possible use with other rules. Note how this differs from a traditional paper-and-pencil checklist, for instance, where the user is confronted with each question regardless of its relevance.

The User Interface: The first versions of REFEREE were written to be used with a terminal connected to a large mainframe computer. In the past year Chavez has transformed the program so as to function at a stand-alone workstation. His first new version was written in an commercial expert system shell (KEE) which rested on an INTERLISP base; however, we then re-wrote the program for the Texas Instrument Explorer in CommonLisp.

The program code is now entirely independent of the knowledge required for reading papers. REFEREE has a new interface that is intuitive and consistent. There is an innovative consultation mode in which questions are presented in free-format menus. The dialogues are mixed-initiative and of mixed levels, allowing the user such options as requesting more detailed questions or cutting off apparently fruitless lines of questioning. With the new REFEREE prototype, the user interacts with the machine using a mouse-pointing device, as with the Macintosh. All questions are asked in a similar format. Finally, the screen enables the user to orient himself at all times, obviating the need for special commands to help the user "navigate" through the knowledge base. Our expert recently provided the best indication of the useability of this new system. After only a brief introduction to the new machine and interface, he was able - for the first time - to run an entire consultation by himself.

Current Status: At this point, REFEREE is a stable prototype that enables the clinician to read clinical trials more critically. As such, REFEREE represents only the first step in a larger research plan, the automation of knowledge acquisition (see section on Research Plans, below). Current work in the restricted domain of clinical trials will, we hope, illustrate general principles in the design of decision makers that gather expertise from written text and multiple knowledge sources.

D. Relevant Publications

Haggerty, J.: *REFEREE and RULECRITIC: Two prototypes for assessing the quality of a medical paper*. REPORT KSL-84-49. Master's Thesis, Stanford University, May 1984.

E. Funding Support

REFEREE currently receives only a small amount of funding. Most of the research is performed in time contributed by the researchers to this project.

Title: Knowledge-Based Systems Research

PI: Edward A. Feigenbaum

Agency: Defense Advanced Projects Research Agency

Grant identification number: N00039-86-0033

Total award period and amount: 10/1/85 - 9/30/88 $4,130,230 (in negotiation) (direct and indirect)

Current award period and amount: 10/1/86 - 9/30/87 $1,549,539 (direct and indirect) REFEREE component is $29,296, or 1.9 % of grant total.

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE

A. Medical Collaborations

Dr. Brown and Dr. Feldman of the Stanford University School of Medicine are actively involved in the REFEREE project and are the primary domain experts for this project.

C. Critique of Resource Management

The SUMEX computer resource and Lisp workstations have been very important for the work to date, and the SUMEX staff has continued to be very cooperative with the REFEREE project.

III. RESEARCH PLANS

A. Goals & Plans

The overall objective of the REFEREE project is to use recent Artificial Intelligence techniques to build a system that helps the informed but statistically non-expert reader to evaluate critically the medical literature on randomized controlled trials (RCT's). This system will contain and be able to apply dynamically the detailed specialized knowledge of Dr. Byron W. Brown, a biostatistician expert in the design and evaluation of randomized controlled trials. We have divided our overall objective into two goals:

- Goal 1 is the construction of an expert system to help readers (e.g. medical students, medical researchers, clinicians, journal editors, or editorial assistants) assess the credibility of a *single* conclusion drawn from a *single* journal report of a randomized controlled trial. We have already made substantial progress toward this goal with the development of the prototype REFEREE system.

- Goal 2 is the expansion of REFEREE to an expert system that can be used by a similar range of readers to facilitate the evaluation of *multiple* reports based on randomized controlled trials. This expanded system, to be known as the REVIEWER, will thus perform meta-analysis.

The task of extending and refining the prototype REFEREE system in order to achieve these goals can be characterized in terms of three dimensions:

- Making the system more accessible to a variety of people by improving the user interface, validating the system's performance with different types of users, and providing an explanatory capability

- Expanding the knowledge base by continuing the knowledge acquisition process to cover additional types of RCT's

- Improving the inference engine to ensure consistency of the knowledge base and to focus the consultation process on questions relevant to the situation and the individual user.

The specific steps that are planned for the enhancement of the REFEREE system include the following:

1. Critique individual clinical trials according to the methodological quality of the trial;

2. Measure the efficacy of treatment as demonstrated in a randomized control trial;

3. Compare and contrast the credibility and efficacy of treatment reported by multiple journal articles; and

4. Combine the *qualitative* techniques of heuristic reasoning and the *quantitative* methods of statistical meta-analysis to extract a consensus opinion from multiple knowledge sources.

In addition, plans for Goal 2, the REVIEWER system to analyze multiple RCT's and form a consensus judgment, include:

1. Complete a review of the available literature on meta-analysis and augment the REFEREE prototype to produce estimators for meta-analysis and incorporate expert knowledge on the appropriateness of these methods.

2. Add explicit and heuristic knowledge needed for the calculation of robust, non-parametric estimators of effect size.

3. Construct a prototype of a system that builds categorical models in the domain of meta-analysis, to perform autonomous investigations in the domain of statistical model-building. The REVIEWER will utilize expert knowledge in biostatistics to guide its search for meaningful models.

4. Build a prototype of a system that can explore the domain of regression models for multiple RCT's that will use expert knowledge in its selection of predictor variables.

5. Package the REVIEWER in a form suitable for use by physicians and their assistants.

6. Verify the expertise of the REVIEWER system on a suite of papers drawn from clinical trials, similar to the validation of REFEREE above.

B. Justification for continued SUMEX use

The local area network maintained by the SUMEX staff is essential to the effective development and use of the REFEREE system on Lisp workstations. The availability of the Xerox workstations makes possible the evaluation of prototypes in that environment, and also facilitates the development of good user interfaces. The connections through the 2060 to local and national computer networks such as ARPAnet are important for sharing ideas and results with other medical researchers.

C. Need for other computing resources

The REFEREE project needs access to an additional high performance Lisp workstation to assist in the development and execution of the REFEREE programs. Such a machine is important to explore user interface issues, in addition to building the knowledge base for current and planned development. In addition, we intend to explore the implementation of REFEREE on less expensive personal computers such as the Macintosh II and other high performance machines. We anticipate the need for at least two of these machines for transporting our system and developing new modes of interaction with both naive and experienced users.

IV.D. Pilot AIM Projects

Following is a description of the informal pilot project currently using the AIM portion of the SUMEX-AIM resource, pending funding, full review, and authorization.

In addition to the progress report presented here, an abstract is submitted on a separate Scientific Subproject Form.

IV.D.1. PATHFINDER Project

PATHFINDER Project

Bharat Nathwani, M.D.
Department of Pathology
University of Southern California

Lawrence M. Fagan, M.D., Ph.D.
Department of Medicine
Stanford University

I. SUMMARY OF RESEARCH PROGRAM

A. Project Rationale

Our project addresses difficulties in the diagnosis of lymph node pathology. Five studies from cooperative oncology groups have documented that, while experts show agreement with one another, the diagnosis made by practicing pathologists may have to be changed by expert hematopathologists in as many as 50% of the cases. Precise diagnoses are crucial for the determination of optimal treatment. To make the knowledge and diagnostic reasoning capabilities of experts available to the practicing pathologist, we have developed a pilot computer-based diagnostic program called PATHFINDER. The project is a collaborative effort of the University of Southern California and the Stanford University Medical Computer Science Group. A pilot version of the program provides diagnostic advice on 72 common benign and malignant diseases of the lymph node based on 110 histologic features. Our research plans are to develop a full-scale version of the computer program by substantially increasing the quantity and quality of knowledge and to develop techniques for knowledge representation and manipulation appropriate to this application area. The design of the program has been strongly influenced by the INTERNIST/CADUCEUS program developed on the SUMEX resource.

PATHFINDER computer science research is focused on the exploration and extension of formal techniques for decision making under uncertainty. Research foci include (1) the assessment and representation of important probabilistic dependencies among morphologic features and diseases, (2) the representation of knowledge about the progression of disease over time, (3) the acquisition and use of independent expert knowledge bases, (4) the customization of the system's reasoning and explanation behaviors to reflect the expertise of the user, and, (5) the explanation of complex formal reasoning techniques.

Toward the pragmatic goal of constructing a useful pathology teaching and decision support system, PATHFINDER investigators are attempting to use intelligent computation to substantially increase the quantity and quality of pathology knowledge available to pathologists. Important areas of this knowledge integration task involve ongoing research on the crisp definition important morphologic features and feature severities, the synthesis of information from multiple experts, the translation among multiple pathology classification schemes, and the incorporation of knowledge about advances in immunology, cytogenetics, cell kinetics, and immunogenetics.

A group of expert pathologists from several centers in the U.S. have showed interest in the program and helped to provide the structure of the knowledge base for the PATHFINDER system.

B. Medical Relevance and Collaboration

One of the most difficult areas in surgical pathology is the microscopic interpretation of lymph node biopsies. Most pathologists have difficulty in accurately classifying lymphomas. Several cooperative oncology group studies have documented that while experts show agreement with one another, the diagnosis rendered by a "local" pathologist may have to be changed by expert lymph node pathologists (expert hematopathologists) in as many as 50% of the cases.

The National Cancer Institute recognized this problem in 1968 and created the Lymphoma Task Force which is now identified as the Repository Center and the Pathology Panel for Lymphoma Clinical Studies. The main function of this expert panel of pathologists is to confirm the diagnosis of the "local" pathologists and to ensure that the pathologic diagnosis is made uniform from one center to another so that the comparative results of clinical therapeutic trials on lymphoma patients are valid. An expert panel approach is only a partial answer to this problem. The panel is useful in only a small percentage (3%) of cases; the Pathology Panel annually reviews only 1,000 cases whereas more than 30,000 new cases of lymphomas are reported each year. A panel approach to diagnosis is not practical and lymph node pathology cannot be routinely practiced in this manner.

We believe that practicing pathologists do not see enough case material to maintain a high level of diagnostic accuracy. The disparity between the experience of expert hematopathology teams and those in community hospitals is striking. An experienced hematopathology team may review thousands of cases per year. In contrast, in a community hospital, an average of only ten new cases of malignant lymphomas are diagnosed each year. Even in a university hospital, only approximately 100 new patients are diagnosed every year.

Because of the limited numbers of cases seen, pathologists may not be conversant with the differential diagnoses consistent with each of the histologic features of the lymph node; they may lack familiarity with the complete spectrum of the histologic findings associated with a wide range of diseases. In addition, pathologists may be unable to fully comprehend the conflicting concepts and terminology of the different classifications of non-Hodgkin's lymphomas, and may not be cognizant of the significance of the immunologic, cell kinetic, cytogenetic, and immunogenetic data associated with each of the subtypes of the non-Hodgkin's lymphomas.

In order to promote the accuracy of the knowledge base development we will have participants for multiple institutions collaborating on the project. Dr. Nathwani will be joined by experts from Stanford (Dr. Dorfman), St. Jude's Children's Research Center -- Memphis (Dr. Berard) and City of Hope (Dr. Burke).

C. Highlights of Research Progress

C.1 Previous Accomplishments

Since the project's inception in September, 1983, we have constructed several versions of PATHFINDER. The first several versions of the program were *rule-based* systems like MYCIN and ONCOCIN which were developed earlier by the Stanford group. We soon discovered, however, that the large number of overlapping features in diseases of the lymph node would make a rule-based system cumbersome to implement. We next considered the construction of a *hybrid system*, consisting of a rule-based algorithm that would pass control to an INTERNIST-like scoring algorithm if it could not confirm the existence of classical sets of features. We finally decided that a modified form of the INTERNIST program would be most appropriate. The original version of PATHFINDER is written in the computer language Maclisp and runs on the SUMEX DEC-20. This was transferred to Portable Standard Lisp (PSL) on the DEC-20, and

later transferred to PSL on the HP 9836 workstations. Two graduate students, David Heckerman and Eric Horvitz, designed and implemented the program and are continuing to lead research on the project.

The prototype knowledge base was constructed by Dr. Nathwani. During the early part of 1984, we organized two meetings of the entire team, including the pathology experts, to define the selection of diseases to be included in the system, and the choice of features to be used in the scoring process.

During the last two years, we have focused on methodologies for more accurately representing expert beliefs. In particular, we have used *influence diagrams* to represent dependencies among features in the PATHFINDER knowledge base. A great deal of effort has been devoted to assessing and representing the intricate relationships among features that exist in the domain. We believe that this process will help to overcome some of the limitations of medical diagnostic systems.

We have also focused on the problem of complex information-theoretic inference. The explanation of a systems diagnostic behavior has been found to be of extreme importance to physicians. Unfortunately, it is often difficult to explain reasoning based on optimal models of inference. We have worked on the use of a set of alternative abstraction hierarchies to control inference. Our current techniques enable us to trade off optimality for the transparency of reasoning. We are now studying the control of this tradeoff to optimize inference.

C.1 The PATHFINDER knowledge base

The basic building block of the PATHFINDER knowledge base is the disease profile or *frame*. Each disease frame consists of *features* useful for diagnosis of lymph node diseases. Currently these features include histopathologic findings seen in both low- and high-power magnifications. Each feature is associated with a list of exhaustive and mutually exclusive *values*. For example, the feature *pseudofollicularity* can take on any one of the values *absent*, *slight*, *moderate*, or *prominent*. These lists of values give the program access to *severity* information. In addition, these lists eliminate obvious interdependencies among the values for a given feature. For example, if pseudofollicularity is *moderate*, it cannot also be *absent*.

Qualitative dependencies among features for each disease are represented using the influence diagram methodology mentioned above. An influence diagram contains *nodes* and *arcs*. Nodes represent features and arcs represent dependencies among features. In particular, an arc is drawn from one feature to another when an expert believes that knowing one feature can change his beliefs that another feature will take on its possible values even when the diagnosis is known. Probabilities are used to quantitate the beliefs asserted by the expert.

C.2 Hewlett-Packard Workstation

Through the USC-affiliated Information Sciences Institute, Dr. Nathwani has obtained a Hewlett-Packard Workstation that is similar to the 9836. The Pathfinder program has been brought up on this machine. This means that the program now exists on three different machines, in three separate locations, using one standard language (Portable Standard Lisp). Thus, the need for support of networked machines and communications has increased during this last year. Current plans are to move the system onto the Macintosh II system.

D. Publications Since January 1984

1. Horvitz, E.J., Heckerman, D.E., Nathwani, B.N. and Fagan, L.M.: *Diagnostic Strategies in the Hypothesis-directed PATHFINDER System, Node Pathology*. HPP Memo 84-13. Proceedings of the First Conference on Artificial Intelligence Applications, Denver, Colorado, Dec., 1984.

2. Heckerman, D. E., and Horvitz, E. J., "The Myth of Modularity in Rule-based Systems," in Uncertainty in Artificial Intelligence, Vol. 2, J. Lemmer, L. Kanal, ed., North Holland, New York, 1987.

3. Horvitz, E.J., Heckerman, D.E., Nathwani, B.N. and Fagan, L.M.: *The Use of a Heuristic Problem-solving Hierarchy to Facilitate the Explanation of Hypothesis-directed Reasoning*. KSL Memo 86-2. Proceedings of MedInfo, Washington D.C., October, 1986.

4. Horvitz, E. J., "Toward a Science of Expert Systems," Invited Paper, Computer Science and Statistics: Proceedings of the 18th Symposium on the Interface, American Statistical Association, March, 1986, pgs. 45-52.

5. Heckerman, D.E., "An Axiomatic Framework for Belief Updates," in Uncertainty in Artificial Intelligence, Vol. 2, J. Lemmer, L. Kanal, ed., North Holland, New York, 1987.

E. Funding Support

Research Grant submitted to National Institutes of Health
Grant Title: "Computer-aided Diagnosis of Malignant Lymph Node Diseases"
Principal Investigator: Bharat Nathwani
Funding for three years from the National Library of Medicine
1 RO1 LM 04529
$766,053 (direct and indirect)

Professional Staff Association, Los Angeles County Hospital, $10,000.

University of Southern California, Comprehensive Cancer Center, $30,000.

Project Socrates, Univ. of Southern Calif., Gift from IBM of IBM PC/XT.

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE

A. Medical Collaborations and Program Dissemination via SUMEX

Because our team of experts are in different parts of the country and the computer scientists are not located at the USC, we envision a tremendous use of SUMEX for communication, demonstration of programs, and remote modification of the knowledge base. The proposal mentioned above was developed using the communication facilities of SUMEX.

B. Sharing and Interaction with Other SUMEX-AIM Projects

Our project depends heavily on the techniques developed by the INTERNIST/CADUCEUS project. We have been in electronic contact and have met with members of the INTERNIST/CADUCEUS project, as well as been able to utilize information and experience with the INTERNIST program gathered over the years through the AIM conferences and on-line interaction. Our experience with the

extensive development of the pathology knowledge base utilizing multiple experts should provide for intense and helpful discussions between our two projects.

The SUMEX pilot project, RXDX, designed to assist in the diagnosis of psychiatric disorders, is currently using a version of the PATHFINDER program on the DEC-20 for the development of early prototypes of future systems.

C. Critique of Resource Management

The SUMEX resource has provided an excellent basis for the development of a pilot project. The availability of a pre-existing facility with appropriate computer languages, communication facilities (especially the TYMNET network), and document preparation facilities allowed us to make good progress in a short period of time. The management has been very useful in assisting with our needs during the start of this project.

III. RESEARCH PLANS

A. Project Goals and Plans

Collection and refinement of knowledge about lymph node pathology

The knowledge base of the program is about to undergo revision by the experts, and then will be extensively tested. A logical next step would be to extend the program to clinical settings, as well as possible extensions of the knowledge base.

Other possible extensions include: developing techniques for simplifying the acquisition and verification of knowledge from experts, and creating mapping schemes that will facilitate the understanding of the many classifications of non-Hodgkin's lymphomas. We will also attempt to represent knowledge about special diagnostic entities, such as multiple discordant histologies and atypical proliferations, which do not fit into the classification methods we have utilized.

Representation Research

We hope to enhance the INTERNIST-1 model by structuring features so that overlapping features are not incorrectly weighted in the decision making process, implementing new methods for scoring hypotheses, and creating appropriate explanation capabilities.

B. Requirements for Continued SUMEX Use

We are currently dependent on the SUMEX computer for the use of the program by remote users, and for project coordination. We have transferred the program over to Portable Standard Lisp which is used by several users on the SUMEX system. While the switch to workstations has lessened our requirements for computer time for the development of the algorithms, we will continue to need the SUMEX facility for the interaction with each of the research locations specified in our NIH proposal. The HP equipment is currently unable to allow remote access, and thus the program will have to be maintained on the 2060 for use by all non-Stanford users.

C. Requirements for Additional Computing Resources

Most of our computing resources will be met by the 2060 plus the use of the Macintosh II workstations. We will need additional file space on the 2060 as we quadruple the size of our knowledge base through the construction of multiple knowledge bases. We will continue to require access to the 2060 for communication purposes, access to other programs, and for file storage and archiving.

D. Recommendations for Future Community and Resource Development

We encourage the continued exploration by SUMEX of the interconnection of workstations within the mainframe computer setting. We will need to be able to quickly move a program from workstation to workstation, or from workstation back and forth to the mainframe. Software tools that would help the transfer of programs from one type of workstation to another would also be quite useful. Until the type of workstations that we are using in this research becomes inexpensive, we will continue to need a machine like SUMEX to provide others with a chance to experiment with our software.

IV.D.2. RXDX Project

RXDX Project

Robert Lindsay, Ph.D.
Michael Feinberg, M.D., Ph.D.
University of Michigan
Ann Arbor, Michigan

I. SUMMARY OF RESEARCH PROGRAM

A. Project Rationale

We are developing a prototype expert system that could act as a consultant in the diagnosis and management of depression. Health professionals will interact with the program as they might with a human consultant, describing the patient, receiving advice, and asking the consultant about the rationale for each recommendation. The program uses a knowledge base constructed by encoding the clinical expertise of a skilled psychiatrist in a set of rules and other knowledge structures. It will use this knowledge base to decide on the most likely diagnosis (endogenous or nonendogenous depression), assess the need for hospitalization, and recommend specific somatic treatments when this is indicated (e.g., tricyclic antidepressants). The treatment recommendation will take into account the patient's diagnosis, age, concurrent illnesses, and concurrent treatments (drug interactions).

B. Medical Relevance and Collaboration

There is a documented shortage of psychiatrists in the US (GMENAC, 1980), and the estimates of the prevalence of psychiatric illness used to develop that report were lower than the figures in recent population surveys (Myers et al., 1984). Further, most prescriptions for antidepressants are written by non-psychiatrists (Johnson, 1974; Kline, 1974) and the great majority of depressed patients seen by a sample of primary care physicians were treated inappropriately (Weissman et al., 1981). These data highlight the need for improving the treatment provided to the majority of mentally ill patients. We believe that computers can act as consultants to non-psychiatrist clinicians, resulting in improved patient care.

The potential benefits to psychiatry include: making relatively skilled psychiatric consultation widely available in underserved areas, including some public mental health facilities where patients are seen by non-psychiatrists and have relatively little direct patient-physician contact; providing non-psychiatrically trained physicians with additional information about psychiatric diagnosis and treatment; avoiding errors of oversight caused by inaccessible patient data; and increased productivity in patient care. Like any good consultant, the program will be able to teach the interested user, and can function as a teaching tool independent of direct clinical application.

C. Highlights of Research Progress

Our major project during the past year has been an expert system for the somatic treatment of (endogenous) depression, where somatic treatment includes antidepressant drugs, electroshock, and lithium. We are writing this system using KEE, an expert system shell generously donated by Intellicorp, running on a Xerox 1108 workstation. We have been able to incorporate the work we did earlier on SUMEX, either directly